Chapter 313

Passing-Bablok Regression for Method Comparison

Introduction

Passing-Bablok regression for method comparison is a robust, nonparametric method for fitting a straight line to two-dimensional data where both variables, X and Y, are measured with error. It is useful when you have two devices that should give the same measurements and you want to compare them. This is accomplished by estimating a linear regression line and testing whether the intercept is zero and the slope is one. The Passing-Bablok regression procedure fits the intercept (β_0) and the slope (β_1) of the linear equation

$$Y = \beta_0 + \beta_1 X.$$

The estimate of the slope (B_1) is calculated as the median of all slopes that can be formed from all possible pairs of data points, except those pairs that result in a slope of 0/0 or -1. To correct for estimation bias caused by the lack of independence of these slopes, the median is *shifted* by a factor K which is the number of slopes that are less than -1. This creates an approximately unbiased estimator. The estimate of the intercept (B_0) is the median of $Y_i - B_1 X_i$.

The intercept is interpreted as the systematic bias (difference) between the two methods. The slope measures the amount of proportional bias (difference) between the two methods.

Experimental Design

Typical designs suitable for Passing-Bablok regression include up to 3000 paired measurements, (x_i, y_i), $i = 1, \ldots, n$, similar to the common input for simple linear regression. Typical data of this type are shown in the table below.

Typical Data for Passing-Bablok Regression

Subject	X	Y
1	7	7.9
2	8.3	8.2
3	10.5	9.6
4	9	9
5	5.1	6.5
6	8.2	7.3
7	10.2	10.2
8	10.3	10.6

Technical Details

The methods and results in this chapter are based on the formulas given in Passing and Bablok (1983, 1984) and Bablok, Passing, Bender, and Schneider (1988).

Assumptions

Passing-Bablok regression requires the following assumptions:

1. The variables *X* and *Y* are highly, positively correlated (for method comparison only).

2. The relationship between *X* and *Y* is linear (straight-line).

3. No special assumptions are made about the distributions (including the variances) of *X* and *Y*.

Passing-Bablok Estimation

Define X_i and Y_i, *i* = 1, ... , *n*, as the values for two variables, each sampled with error to give the observed values x_i and y_i, respectively. For each of the $N = \binom{n}{2}$ possible pairs of points, define the slope by

$$S_{ij} = \frac{y_i - y_j}{x_i - x_j}$$

If $x_i = x_j$ and $y_i = y_j$, the result is 0/0 which is undefined. Omit these pairs from the calculation of the slope.

If $x_i = x_j$ and $y_i > y_j$, the result is ∞. Enter a large positive number. Since we are using the median of the S_{ij}, the value won't matter.

If $x_i = x_j$ and $y_i < y_j$, the result is −∞. Enter a large negative number. Since we are using the median of the S_{ij}, the value won't matter.

If $S_{ij} = -1$, omit the pair from the calculation of the slope. This corrects for the bias caused by the lack of independence in the S_{ij}.

If $S_{ij} < -1$, use the pair and add one to a counter *K*. This value will be used to shift the median.

The slope B_1 is the shifted median of S_{ij}, where the median is shifted to the right *K* steps.

Using this slope, calculate intercept B_0 as the median of all *n* of the quantities $y_i - B1x_i$.

Confidence Bounds

Calculate the confidence bounds for β_1 as follows. Let $z_{\alpha/2}$ be the 1 – α/2 quantile of the standard normal distribution. Let

$$C_{\alpha/2} = z_{\alpha/2} \sqrt{\frac{n(n-1)(2n+5)}{18}}$$

Now calculate

$$M_1 = \left[\frac{N - C_{\alpha/2}}{2} \right]$$

Here, [U] rounds U to the nearest integer.

Also calculate

$$M_2 = N - M_1 + 1$$

Finally, a confidence interval for β_1 is given by

$$S_{(M_1 + K)} \le \beta_1 \le S_{(M_2 + K)}$$

Where the S_{ij}'s are sorted.

Confidence limits for the intercept are calculated as follows. Let B_{1L} and B_{1U} be the lower and upper limits for the slope from the last calculation. Now calculate the limits for the intercept as

$$B_{0L} = \text{median}\{y_i - B_{1U} x_i\}$$

and

$$B_{0U} = \text{median}\{y_i - B_{1L} x_i\}$$

Cusum Test of the Linearity Assumption

An important assumption is that the relationship between X and Y is linear. Passing and Bablok (1983) proposed testing this assumption using a modified cusum test. This test proceeds as follows.

Let n_{pos} represent the number of observations with $y_i > B_0 + B_1 x_i$. Similarly, let n_{neg} represent the number of observations with $y_i < B_0 + B_1 x_i$.

Step 1 Assign score r_i to each point as follows.

$$\text{If } y_i > B_0 + B_1 x_i \quad \text{then } r_i = \sqrt{n_{neg}/n_{pos}}$$

$$\text{If } y_i < B_0 + B_1 x_i \quad \text{then } r_i = -\sqrt{n_{pos}/n_{neg}}$$

$$\text{If } y_i = B_0 + B_1 x_i \quad \text{then } r_i = 0$$

Step 2 Assign the distance scores D_i to each observation as follows.

$$D_i = \frac{y_i + \frac{x_i}{B1} - B0}{\sqrt{1 + 1/B1^2}}$$

Step 3 Sort the r_i in the order of the D_i. Label these sorted values $r_{(i)}$.

Step 4 Create the $cusum_{(i)}$ statistics using the following.

$$cusum_{(i)} = \sum_{k=1}^{i} r_{(k)}$$

Step 5 Find the maximum absolute value of $cusum_{(i)}$ across all values of i. Call this $MaxCusum$.

Step 6 The null hypothesis of random arrangement of residuals and therefore a linear relationship between X and Y is tested using the test statistic

$$H = \frac{MaxCusum}{\sqrt{n_{neg} + 1}}$$

This statistic is compared to the Kolmogorov-Smirnov distribution. Details are given in Passing and Bablok (1983). For example, the critical value for a 5% test is 1.36 and for a 1% test is 1.63. Note that we created an approximating polynomial that allows us to calculate p-values for any value of H.

Kendall's Tau Test of the High Correlation Assumption

Passing and Bablok (1983) recommended that a preliminary two-sided test be conducted to determine if Kendall's tau correlation between X and Y is significantly different from zero. They also indicate that this correlation must be positive. Kendall's tau correlation is well documented in the *Correlation* procedure.

Example 1 – Passing-Bablok Regression for Method Comparison

This section presents an example of how to run a Passing-Bablok regression method comparison of the data in the *PassBablok 1* dataset. This dataset contains measurements from two measurement methods on each of 30 items. The goal is to determine if the two measurement systems obtain equivalent measurements.

Setup

To run this example, complete the following steps:

1   Open the PassBablok 1 example dataset
- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **PassBablok 1** and click **OK**.

2   Specify the Passing-Bablok Regression for Method Comparison procedure options
- Find and open the **Passing-Bablok Regression for Method Comparison** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Y Variable **Method2**
X Variable **Method1**

Reports Tab

Residuals..**Checked**

Plots Tab

Passing-Bablok Regression Scatter Plot ....**Checked**
Residual Plot...**Checked**

3   Run the procedure
- Click the **Run** button to perform the calculations and generate the output.

Run Summary

Run Summary

Item	Value	Item	Value
Y Variable	Method2	Rows Used	30
X Variable	Method1	B0: Intercept	-0.0092
		B1: Slope	0.9986

This report gives a summary of the input and various descriptive measures about the Passing-Bablok regression.

Descriptive Statistics

Descriptive Statistics

Item	Y	X	Y - X
Variable	Method2	Method1	Difference
N	30	30	30
Mean	62.99333	63.14	-0.1466667
Std Dev	23.5886	23.6153	2.2578
Std Error	4.3067	4.3115	0.4122
COV	0.3745	0.3740	-15.3942
Minimum	12.7	14.2	-3.8
First Quartile	47.425	48.8	-0.5
Median	64.15	61.4	-0.1
Third Quartile	83.075	84.1	0.3
Maximum	98.6	99.1	10.2

This report gives descriptive statistics about the variables used in the regression.

Kendall's Tau Correlation Confidence Interval and Hypothesis Test

Kendall's Tau Correlation Confidence Interval and Hypothesis Test

Kendall's Tau Correlation	Lower 95%: Conf. Limit of Tau	Upper 95%: Conf. Limit of Tau	Z-Value for Testing H0: $\rho = 0$	P-Value	Reject H0 that $\rho = 0$ at $\alpha = 0.05$?
0.9724	0.9546	0.9833	7.5289	0.0000	Yes

This procedure assumes that Kendall's Tau is significantly greater than zero.
This assumption cannot be rejected.

The section reports an analysis of Kendall's tau correlation. It provides both a confidence interval and a significance test. The main thing to look for here is that the correlation (0.9724 in this example) is large and positive. This can be surmised from both the confidence interval and the p-value.

Cusum Test of Linearity

Cusum Test of Linearity

Maximum Cusum	Linearity Z Value	Linearity P-Value	Reject Linearity at α = 0.05?
3.0000	0.7500	> 0.3	No

This procedure assumes that only a positive linear relationship exists between the two measurements.
This assumption cannot be rejected.

The section reports the results of the cusum test of linearity. When linearity is rejected, the assumption that the data supports this assumption is rejected and the analysis should not be used.

Regression Coefficients

Regression Coefficients

Item	Intercept B0	Slope B1
Regression Coefficient	-0.0092	0.9986
Lower 95% Conf. Limit of β(i)	-0.7486	0.9856
Upper 95% Conf. Limit of β(i)	0.7987	1.0099

Method Comparison Hypothesis Test
Hypothesis being tested, H0: β0.LCL < 0 < β0.UCL and β1.LCL < 1 < β1.UCL.
Cannot reject H0. There is not enough evidence to conclude that the methods are not equal.

This section reports the regression coefficients, along with their analytic confidence limits. The conclusion is reported on the last line. **This is the main conclusion desired of this report.**
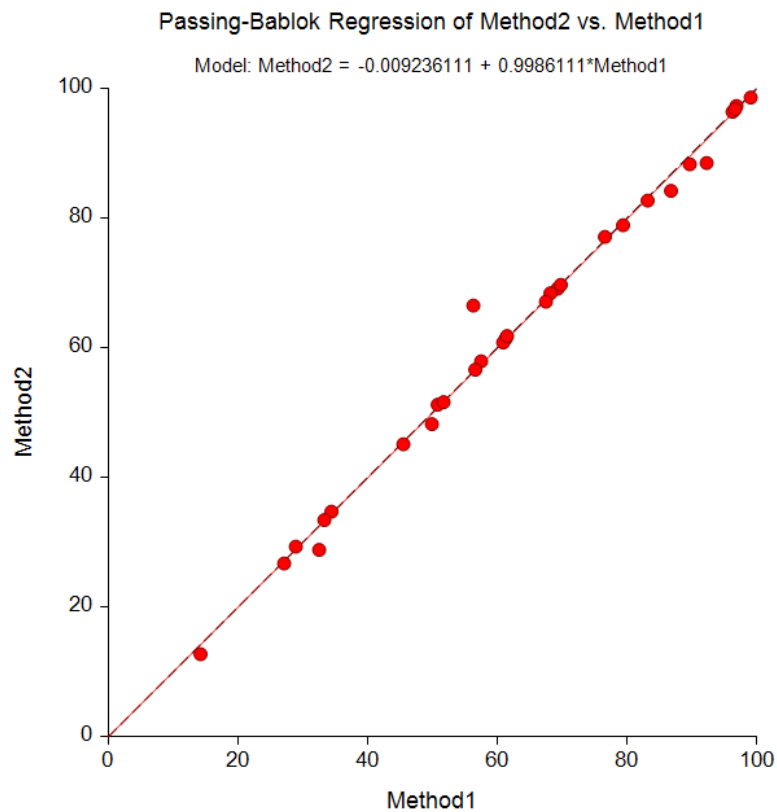
Residuals

Residuals

Row	X	Y	Difference Y - X	Predicted Yhat	Residual (Y-Yhat)
1	69.3	69.1	-0.2000	69.1945	0.0945
2	27.1	26.7	-0.4000	27.0531	0.3531
3	61.3	61.4	0.1000	61.2056	-0.1944
4	50.8	51.2	0.4000	50.7202	-0.4798
5	34.4	34.7	0.3000	34.3430	-0.3570
6	92.3	88.5	-3.8000	92.1626	3.6626
7	57.5	57.9	0.4000	57.4109	-0.4891
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

This section reports the residuals and the difference for each of the input data points.

Passing-Bablok Regression Scatter Plot

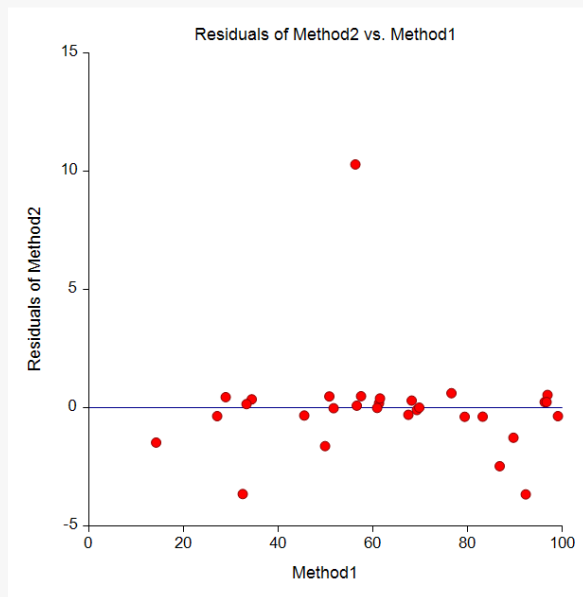**Passing-Bablok Regression Scatter Plot**
_____

Passing-Bablok Regression of Method2 vs. Method1

Model: Method2 = -0.009236111 + 0.9986111*Method1

This report shows both the 45-degree line (Y = X) and the fitted Passing-Bablok regression line. In this example, the regression line hides the 45-degree line.

Residual Plot

Residual Plots
_____

This plot emphasizes the deviation of the points from the regression line.